



SMALL WONDERS

The biggest advances in computing now rely on the tiniest resources. Mike Bedford considers how shrinking the processor could make computers faster than ever before

The microprocessor has given us computers that sit on our desk rather than fill an entire room. But have you ever wondered why semiconductor manufacturers are still so keen to reduce a chip's feature size? The processor accounts for a small fraction of the space inside your PC's case. Would it be such a problem if the feature size was 90 nanometres, 90nm (90 billionths of a millimetre) or 65nm? Not really. At the moment, shaving off a few nanometers doesn't make a difference to how the processor fits in the system case. But that is not the main reason why processors need to get ever smaller.

Moore's Law, coined in 1965 by Intel pioneer Gordon Moore, states that the number of transistors on a chip doubles every 18 months (although this was later amended to every two years). The result has been a continual improvement in microprocessor performance. Today those extra transistors are being used to give us processors with two or even four cores. If this trend continues, we'll have 128-core chips within a decade and, without a corresponding reduction in the feature size, that would be a seriously large chip.

The laws of physics state that a fast chip has to be a small chip, as explained in the box on page 148. Experts believe that the traditional

methods of manufacturing chips can't be scaled down much further. But as one technology starts to run out of steam, some researchers believe they have a replacement. Using nanotechnology, which deals with materials at a molecular and atomic level, they've come up with a radical new way of making electronic circuits.

SEMICONDUCTOR MANUFACTURING

Since the first commercial integrated circuits of the early 1960s, all chips have been made using a technique called photolithography. This is a multi-stage process in which the circuit is built up in layers. Layers include conductive metal to provide the interconnections, silicon oxide to act as an insulator, and polycrystalline silicon, which is doped to give the semiconductor properties needed to form transistors. Each of these layers has to be formed into an intricate pattern to create the required electronic circuit.

To do that, when each layer is first applied, it is given a coating of so-called photo-resist, on to which the required pattern is projected optically. The photo-resist is then developed, like a photographic film, so only those portions of the photo-resist that correspond to the projected pattern remain. The chip is immersed in a liquid called an etchant, which dissolves those areas that were not protected by the

photo-resist. Finally, the remaining photo-resist is removed.

At first sight, it seems as if this method could be used for ever-smaller features by projecting the patterns at an ever-smaller scale. But that's forgetting about the properties of light. As the feature size approaches the wavelength of the light being used, it's no longer possible to focus the pattern sharply on the photo-resist. To date this limitation has been overcome by using light with an ever-shorter wavelength, but since today's state-of-the-art technology uses extreme ultraviolet the technique is running out of steam. Next in line after extreme ultraviolet are electron beams, but to say that this technology is in its infancy would be a gross understatement. Some experts question whether or not it will ever lend itself to large-scale manufacturing. A better method is needed.

The process of photolithography is like trying to carve a pocket watch out of a solid block of steel. A watchmaker would never do that. Instead, the watch is built up from its component parts. Many researchers believe that this is how the chips of the future are going to be made. This up-and-coming field of nanotechnology, and the technique of building a chip from its constituent parts, is referred to as a bottom-up approach. This is in contrast



Fine figures Putting nanotechnology in perspective

In this article we deal with lots of very small figures, figures that are very hard to imagine. So just how small are the features of today's microprocessors, and how much smaller are the nanoscale circuits we've been looking at? Let's put some of these figures into context.

The first-ever electronic switch, the triode valve, was patented 100 years ago by Lee De Forest. More recent valves, which were in widespread use until the 1960s and 1970s, were around 100mm tall. The transistor was invented at Bell Labs in 1947. Discrete transistors (transistors that are not part of an integrated circuit) were about 10mm tall, 10 times smaller than a valve.

In 1958, Jack Kirby of Texas Instruments produced the first integrated circuit (IC). It was an oscillator containing a single

transistor and a handful of other components. At its minimum feature size, it was about 1mm, another tenfold reduction.

WHAT'S INSIDE?

Intel launched the world's first microprocessor in 1971 with a 0.01mm (10 micron) feature size. By 1989, the 486 processor had reached another milestone: the 1 micron feature size. The next milestone, the first sub-0.1 micron feature size, was reached in 2003 with the 90nm (0.09 micron) process.

While we might already be talking in nanometers, this isn't in the realm of true nanotechnology. However, when we move to Franz Himpel's atomic memory, we find a feature size that is another 100 times smaller. The bits in his memory are spaced at less than 1nm.

to the top-down method of today's semiconductor manufacturing.

HISTORY LESSON

It is possible that nanotechnology might one day give us chips with unimaginably small feature sizes. In order to see how, we need to go back to the beginning and see how it all started.

In 1989, Don Eigler and Erhard Schweizer of IBM's Almaden Research Center, California, employed a scanning tunnelling microscope (STM) in a rather unusual way. An STM is a development of the electron microscope that's capable of seeing individual atoms. Instead of using the STM to produce an atomic scale image of a minute object, they found it was possible to press it into service for moving individual atoms around. By cooling a crystal of nickel down to a temperature of 4°K (just above absolute zero), Eigler and Schweizer were able to manipulate 35 atoms of xenon to spell out the letters 'IBM'. This feat took 22 hours, which is around 40,000 times slower than a laser printer could print a whole page of IBMs. But it was a development some pundits believe will change the world.

By 1996, scientists at IBM's Zurich Research Laboratory were working at room temperature and had created the first molecular computing device. Comprising 110 Buckminsterfullerene molecules (a form of carbon containing 60 atoms and known colloquially as a buckyball) that were constrained to move in rows by one-atom tall ridges in the underlying copper

surface, this structure took the form of a minute abacus. Although the Zurich abacus had just 11 rows, allowing it to work on 11-digit numbers, hundreds of rows of these buckyball molecules could fit into the minimum feature size of today's microprocessor chips. However, as with positioning those 35 atoms to create the letters IBM, manipulating this minuscule abacus was a tedious process. "The tool we use [the STM probe] is the equivalent of operating a normal abacus with the Eiffel Tower," explained IBM researcher Dr James Gimzewski.

MOLECULAR VISION

The concept of nanotechnology predated IBM's research efforts by 30 years. In 1959, the Nobel Prize-winning physicist Richard Feynman addressed the American Physical Society with a talk entitled 'There's Plenty of Room at the Bottom'. Feynman painted a picture of a world in which manufacturing involved moving matter around, atom by atom and molecule by molecule.

Those early IBM researchers achieved this feat using a clumsy and painfully slow tool – the scanning tunnel microscope – but Feynman had a more elegant solution. If large tools are too cumbersome to push around atomic particles, why not use smaller tools? Feynman felt the secret to creating molecular-scale tools was through repeated miniaturisation. If a tool could be used to work on objects a tenth of its own size, it could be used to build a tenth-scale replica of itself. This tiny replica could then build

a tool a hundredth of the size of the original and so, within a few generations, we would be down to molecular dimensions.

Feynman's vision was expanded upon by Eric Drexler in his 1986 book *Engines of Creation*. However, Drexler's version of future technology had more sinister overtones. Drexler envisaged intelligent nano-machines called assemblers that would be capable of producing any object the laws of chemistry would allow just by pushing atoms around. They could also create replicas of themselves, so their population could grow exponentially until there was an army of them at your disposal.

Drexler talked about the danger of the 'grey goo' scenario. Slight errors or evolutions that might take place when the assemblers replicate could result in changes to their programmed goals. Errant assemblers might have one goal and one goal only: to survive as a species. The grey goo he referred to was a mass of molecular assemblers, turning the world and everything on it to yet more grey goo. As Drexler pointed out, "The grey goo threat makes one thing clear. We cannot afford certain kinds of accidents with replicating assemblers."

Today nanotechnologists are distancing themselves from this dream. Even Drexler is now talking of different and more benign nanoscale manufacturing techniques. This problem is of particular interest to Ottilia Saxl. He works at the Scottish-based Institute of Nanotechnology, founded in 1997 to raise the awareness of nanotechnology in industry and the public. Saxl explained: "That term [self-replicating molecular assemblers] implies something that is generally thought of as being in the realms of science fiction. But nanotechnology will certainly lead to very clever solutions to what seem like intractable problems today."

While some other experts wouldn't go as far as saying that self-replicating molecular assemblers will never happen, nobody expects to see them used as a manufacturing technique in the foreseeable future. "Given the advances in technology and some imagination, there are often easier ways to get to where we want to be," said Saxl.

NANOTUBE MEMORY

We need to consider these easier methods of nano-manufacturing, and how they are being harnessed to bring us the chips of the future. You were probably told in school chemistry lessons that there are two forms of carbon – graphite and diamond – both of which have

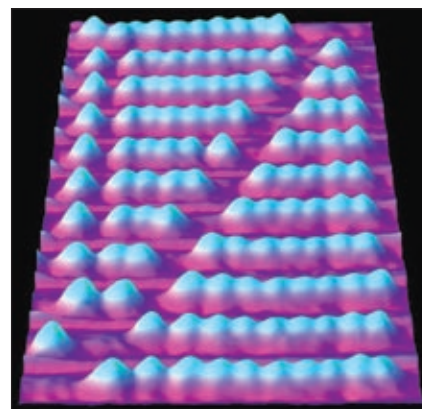
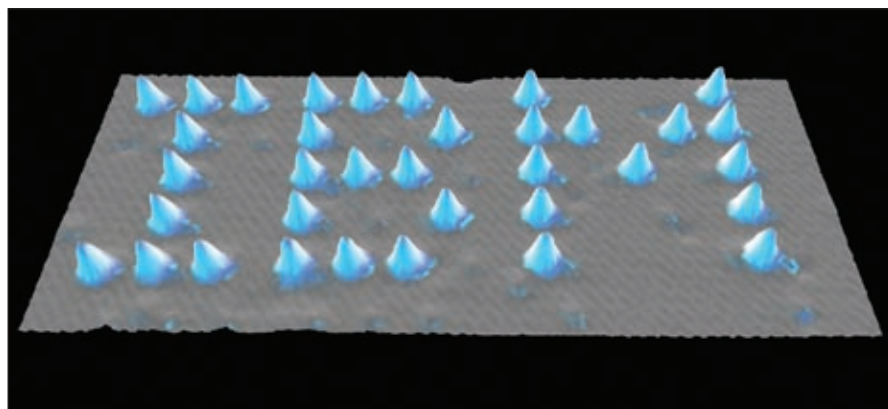
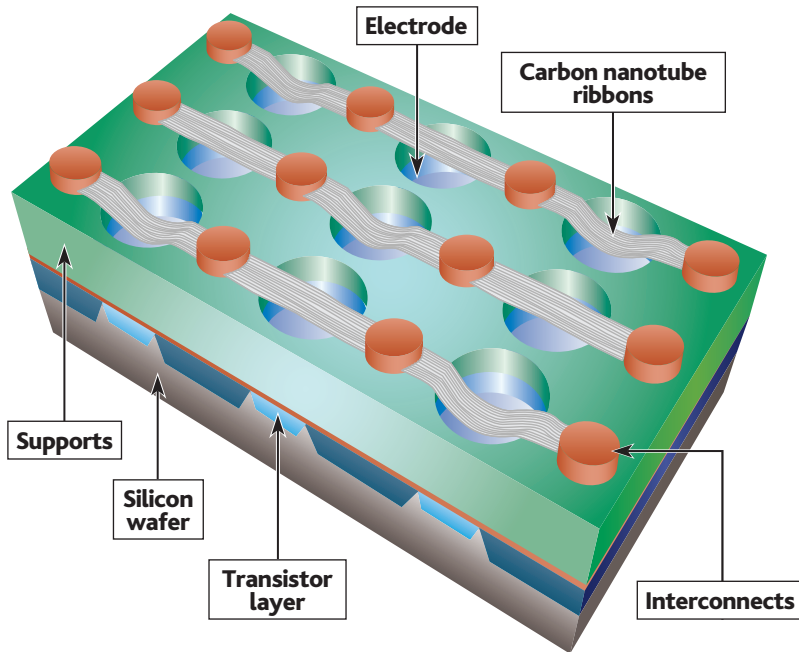


PHOTO CREDIT: IBM

▲ The image above shows the letters IBM, spelt using just 35 atoms. On the right is the first molecular computing device, made up of 110 Buckminsterfullerene molecules held within one-atom-tall ridges. Both pictures were taken using a scanning tunnelling microscope (STM)



▲ Carbon nanotubes could be a replacement for system memory and flash drives. When a nanotube is flexed downwards it makes a circuit and represents a one. If it is in the up position, there is no circuit and it represents a zero. Electrostatic force flexes the carbon nanotube, and this force is generated by the underlying silicon control circuitry. Carbon nanotube memory is as fast as SRAM chips

millions of carbon atoms, but which differ in the way those atoms are joined together. In fact, there are many more allotropes of carbon.

Collectively called fullerenes, these other forms were discovered recently at Rice University in Houston, Texas. We've mentioned one of them already – a molecule comprising 60 carbon atoms arranged in a sphere and called Buckminsterfullerene, which was used to create a nanoscale abacus. Of particular interest to nanotechnologists are carbon nanotubes that are cylindrical fullerenes. The tubes are a few nanometres wide (100,000 times thinner than a human hair) but can be several millimetres long. They have high tensile strength and high electrical conductivity, and they can also form semiconductors. This has led many to predict that they will form the basis of future generations of chips.

A company called Nantero has developed a memory chip using carbon nanotubes. It claims that this memory, called Nano-RAM (or NRAM), combines the best features of dynamic RAM (DRAM), static RAM (SRAM) and flash memory. Paradoxically, these circuits don't use the nanotubes as semiconductors. As Thomas Rueckes, Nantero's chief technology officer and inventor of NRAM, explained, "NRAM uses carbon nanotube nanoelectromechanical switches to represent bits".

Rather than the electronic switches (transistors) we expect to find in chips, the carbon nanotubes act like common light switches. If a nanotube is in the up position, it represents a zero; if it's flexed down to make a circuit, it represents a one. The motive force that is required to flex a carbon nanotube is electrostatic, generated by the underlying silicon

control circuitry. Where carbon nanotube switches differ from ordinary switches is in their size. "The bit state can be changed rapidly since the carbon nanotubes have a very small mass and move only a short distance measured in nanometers," explained Rueckes.

This begs the question, exactly how fast can they operate? "The nanoelectromechanical switches in NRAM can move up and down billions of times per second, which translates to nanosecond switching times. This is similar to the fastest SRAM chips available today, but with the unique added benefit of non-volatility," Rueckes explained. "As the NRAM switches get smaller, they also get faster, so we expect to be able to show even faster speeds in the future."

Nantero's NRAM may use carbon nanotubes, heralded by many as the building blocks of a nano-future, but they are assembled using conventional techniques. The carbon nanotubes are spin-coated on the surface of a silicon wafer and they are etched away using the same photolithographic techniques that are employed to make Pentium chips. "As long as lithography continues to evolve to smaller dimensions, the current NRAM integration will scale along with it, even to sizes below 5nm," said Rueckes.

However, this is a big 'if' given the doubt, as mentioned earlier, over the long-term future of photolithography. "For the post-lithography era, the opportunity exists for self-aligned carbon nanotube circuits that won't be restricted by the fabrication limitations of silicon electronics," said Rueckes. "In this era, carbon nanotube electronics could even completely replace silicon electronics for some applications."

ATOMIC MEMORY

Carbon nanotubes might be small, but they offer far from the ultimate in memory density, as work by Franz Himpel of the University of Wisconsin in Madison has proved. Himpel's goal was to push storage density to the atomic limit and test whether a single atom could be used to store a bit at room temperature. A key question was how closely the atoms could be packed without them interacting.

Quantum leap Making sense of quantum bits

Quantum dots are tiny structures just a few nanometres across that possess some unique properties. To date they've been used to help scientists probe living cells in full colour, but one day they might form the basis of the first practical quantum computer.

An ordinary digital computer manipulates data stored as binary numbers that are composed of a number of bits. Each bit can store a zero or a one. A quantum computer takes advantage of the strange laws of quantum physics, which state that once you get down to the atomic scale, particles can be in two places at once or in two states at the same time. If you use a single atomic or sub-atomic particle as data storage, it can hold both a zero and a one simultaneously.

These quantum bits are called qubits, and the implications are far reaching. Just as a collection of eight ordinary memory bits can hold any number between zero and 255, a collection of eight qubits can store all the numbers between zero and 255 at the same

time. Not only that, but a computer based on this principle – a quantum computer – would be able to carry out calculations on all those numbers at once. The potential increases exponentially with the number of qubits. A collection of 64 qubits could operate on almost two billion billion values simultaneously.

QUBITS MOVEMENT

If qubits come into contact with other particles, they lose their strange properties and start behaving like ordinary bits. This tendency increases with the number of qubits, which explains why, to date, nobody has managed to build a quantum computer with more than just a few qubits. Early quantum computers used qubits that took the form of photons, ions or atoms suspended in magnetic fields, but researchers at Purdue University in Indiana believe that these problems could be overcome by creating qubits on a solid surface.

The technique involves creating cages that can entrap a few electrons on the surface of a

gallium arsenide chip. This pool of electrons is called a quantum bit. By depositing another electron into the pool via one of the chip's gates, all the electrons in the quantum dot will take on the spin of the new electron. The direction of the spin takes one of two values, so it can be used to represent a zero or a one. Manipulating the quantum bit by some of the other gates, the electrons can be put into a state of superposition, which means the electron can have both spins at once.

Don't expect to see a quantum computer on your desk any time soon, though. "What we've done is just a physics experiment," said Albert M Chang, formerly at Purdue and now at Tokyo University. "In future, we'll need to manipulate the spin at very fast rates. But for the moment we have, for the first time, demonstrated the entanglement of two quantum dots, and shown that we can control its properties with great precision. It offers hope that we can reach that future within a decade or so."



Speed trials Capacitance and the faster processor

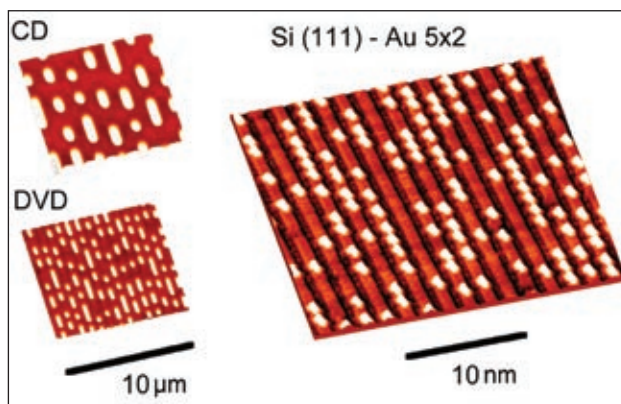
We're all familiar with the struggle to create ever-smaller feature sizes in chips, but this isn't just to provide us with compact processors. It's also essential to give us ever-faster processors.

The most basic building block of any integrated circuit is the transistor. Transistors combine to produce larger building blocks such as AND gates and OR gates and these, in turn, combine to create more complicated building blocks, until we arrive at a complete processor.

The function of a transistor is simple enough: it allows one electrical circuit to be turned on and off by another. A metal-oxide-semiconductor field-effect transistor (MOSFET), the type of transistor used in processors, has three terminals: the gate, the source and the drain. If a sufficiently large voltage is applied between the gate and the source, the transistor will switch on, thereby allowing a current to flow between the source and the drain.

Unfortunately, there's no such thing as a perfect transistor. Real-world transistors have a property called capacitance, which affects their operation. One property of a capacitor is that it can store an electrical charge. Apply a voltage to a capacitor and it will charge up over time, and the larger the capacitance the longer this will take.

Consider what would happen if you applied a voltage between the gate and the source of a real-world transistor with the aim of turning it on. The capacitance between the gate and the source would start to charge up, and only when it reached a sufficient level of charge would the transistor 'see' the necessary threshold voltage for it to turn on. This capacitance, plus other stray capacitances in the transistor, will cause it to have a switching delay. Fortunately, these capacitances can be reduced by reducing the dimensions of the transistor, which is why a small transistor is faster than a large one.



◀ Comparing the scale of Franz Himpsel's atomic memory (right) with CD and DVD

PHOTO CREDIT: FRANZ HIMPSEL

The result of this work was a two-dimensional memory in which the presence or absence of a single atom of silicon was used to represent a one or a zero respectively. Each of the atoms was positioned in the centre of a 4x5 block of atoms, in which the other 19 atoms were used to prevent adjacent bits interacting with each other. Just as the data on a CD or a DVD is arranged in rows, the atoms that Himpsel uses to represent the data are aligned in rows. They are formed naturally by depositing gold on the surface of the silicon at a high temperature. A CD or a DVD has dimensions measured in micrometres, but this molecular memory has features measured in nanometres. "With that density, one could store a million CDs on the area of one CD, enough for the most demanding music collector," Himpsel said.

This sort of memory density sounds impressive, but there are concerns that it might not be a practical proposition. After all, reading from and writing to the memory uses a scanning tunnelling microscope which, as those first-ever nanotechnologists at IBM discovered, isn't the fastest instrument around. Himpsel did admit that for the mass market at the moment "it is totally impractical".

So it's early days yet, but Himpsel thinks the technology could one day have the scope to enter the mainstream. "To date, we've achieved a speed of 100bit/s, but the laws of physics show that it should be possible to increase this to 10Mbit/s. However, that is still a hundred times slower than the speed of hard disks today. Therefore, one would need a hundred reading

heads to reach the speed of hard disks. The IBM lab in Zurich has already developed an array of a thousand read/write heads: the millipede."

SELF-ASSEMBLING CIRCUITS

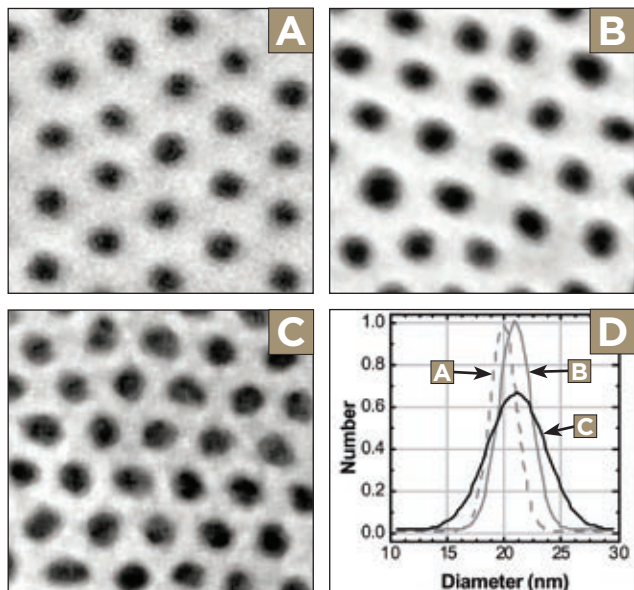
If you looked at the perfectly cubic crystals of the mineral iron pyrite or examined a snowflake under a magnifying glass, you might find it hard to believe that they occur naturally. The laws of chemistry cause many chemical substances to form naturally into regular geometrical structures. Scientists at IBM's Watson Research Center are trying to use this property of certain plastics to cause chips to assemble themselves.

IBM researchers have used a technique called polymer self-assembly to form critical features of a flash memory chip. In the first step, polymer molecules are made to self-assemble into perfect hexagonal arrangements. Next, the polymer material is used as a stencil to reproduce the nanoscale pattern in silicon dioxide, which is more rugged than the polymer and able to withstand higher temperatures. The polymer is then removed. A combination of depositing silicon material and etching is used to leave silicon nanocrystals embedded within the 20 nanometer regions defined originally by the self-assembled polymer. The result is a device with features that are smaller, denser, more precise and more uniform than can be achieved using conventional methods such as lithography.

FROM CHIPS TO CADILLACS

To believe that nanotechnology is just about new ways of assembling electronic circuits would be to underestimate its potential. When he painted a picture of armies of miniscule molecular assemblers, Eric Drexler didn't envisage them creating only such tiny structures as microprocessors. In Drexler's vision, nanotechnology would be used for making anything and everything from cars to jumbo jets to houses, all at knock-down prices. Once that very first assembler had been made, the cost of manufacturing anything would be little more than the cost of the raw materials. Furthermore, even those raw materials would not need to be exotic substances. Because the assemblers would work at the atomic level, they could create esoteric new substances from very basic raw materials.

Today, nanotechnologists have their eyes set on much more down-to-earth applications in medicine and materials science, so talk of a world in which we can have anything at virtually no cost is surely in the realms of science fiction. Still, sometimes it pays to dream; it wouldn't be the first time that science fiction has become science fact. **CS**



◀ IBM's polymer self-assembly. Polymer molecules are made to self-assemble into perfect hexagonal arrangements. The dark areas are 20nm in diameter and spaced 40nm apart (image A). The polymer material is used as a stencil to reproduce the nanoscale pattern in silicon dioxide (image B). A combination of depositing silicon material and etching leaves silicon nanocrystals embedded within the 20nm regions (image C). Image D shows that the dimensions of the hexagonal pattern of the initial polymer (dotted curve) are maintained, as shown by the histogram of the final silicon nanocrystal dimensions (solid curve). The grey curve represents the dimensions of the hexagonal pattern at an intermediate process step

PHOTO CREDIT: IBM